

430 Rec'd PCT/PTO 15 MAR 2000

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## (i) APPLICANT:

- (A) NAME: RHONE-POULENC AGRICULTURE LIMITED
- (B) STREET: FYFIELD ROAD
- (C) CITY: ONGAR
- (D) STATE: ESSEX
- (E) COUNTRY: UNITED KINGDOM
- (F) POSTAL CODE (ZIP): CM5 0HW

(ii) TITLE OF INVENTION: GLUTATHIONE TRANSFERASES

(iii) NUMBER OF SEQUENCES: 19

## (iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0.  
Version #1.30 (EPO)

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1085 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 46..711

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 1..1085
- (D) OTHER INFORMATION: /note= "SEQUENCE OF TaGST1 AND  
ENCODED AMINO ACID SEQUENCE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CAAACACAAG CACAGATCGG TCGAGATTCA AGGCAACCGG GAGCA ATG GCG GGC	54
Met Ala Gly	
GAG AAG GGG CTG GTG CTG CTG GAC TTC TGG GTG AGC CCG TTC GGG CAG	102
Glu Lys Gly Leu Val Leu Leu Asp Phe Trp Val Ser Pro Phe Gly Gln	
CGC GTG CGC ATC GCG CTG GCC GAG AAG GGC CTG CCC TAC GAG TAC GCG	150
Arg Val Arg Ile Ala Leu Ala Glu Lys Gly Leu Pro Tyr Glu Tyr Ala	
GAG GAG GAC CTG ATG GCC GGC AAG AGC GAC CGC CTC CTC CGC GCC AAC	198
Glu Glu Asp Leu Met Ala Gly Lys Ser Asp Arg Leu Leu Arg Ala Asn	
CCG GTG CAT AAG AAG ATC CCG GTG CTC CTC CAC GAC GGC CGT GCC GTC	246
Pro Val His Lys Lys Ile Pro Val Leu Leu His Asp Gly Arg Ala Val	
AAC GAG TCC CTC ATC ATC CTC CAG TAC CTG GAG GAG GCC TTC CCG GAC	294
Asn Glu Ser Leu Ile Ile Leu Gln Tyr Leu Glu Glu Ala Phe Pro Asp	
GCG CCC GCT CTG CTC CCC TCC GAC CCC TAC GCG CGC GCG CAG GCC CGC	342
Ala Pro Ala Leu Leu Pro Ser Asp Pro Tyr Ala Arg Ala Gln Ala Arg	
TTC TGG GCC GAC TAC GTC GAC AAG AAG GTC TAC GAC TGC GGC TCC CGC	390
Phe Trp Ala Asp Tyr Val Asp Lys Lys Val Tyr Asp Cys Gly Ser Arg	
CTC TGG AAG CTC AAG GGC GAG CCG CAG GCG CAG GCG CGC GCC GAG ATG	438
Leu Trp Lys Leu Lys Gly Glu Pro Gln Ala Gln Ala Arg Ala Glu Met	
CTG GAC ATC CTC AAG ACC CTC GAC GGC GCG CTC GGG GAC AAG CCC TTC	486
Leu Asp Ile Leu Lys Thr Leu Asp Gly Ala Leu Gly Asp Lys Pro Phe	
TTC GGC GGC GAC AAG TTC GGG TTC GTC GAC GCC GCC TTC GCG CCC TTC	534
Phe Gly Gly Asp Lys Phe Gly Phe Val Asp Ala Ala Phe Ala Pro Phe	
ACC GCG TGG TTC CAC AGC TAC GAG AGG TAC GGC GAG TTC AGC CTG CCG	582
Thr Ala Trp Phe His Ser Tyr Glu Arg Tyr Gly Glu Phe Ser Leu Pro	
GAG GTG GCG CCC AAG ATC GCC GCG TGG GCC AAG CGC TGC GGC GAG CGG	630
Glu Val Ala Pro Lys Ile Ala Ala Trp Ala Lys Arg Cys Gly Glu Arg	
GAG AGC GTC GCC AAG AGC CTC TAC TCG CCG GAC AAG GTG TAC GAC TTC	678
Glu Ser Val Ala Lys Ser Leu Tyr Ser Pro Asp Lys Val Tyr Asp Phe	
ATC GGC CTG CTC AAG AAG AAG TAC GGC ATC GAG TA GGCGCGCCGA	723
Ile Gly Leu Leu Lys Lys Lys Tyr Gly Ile Glu	

CGGACGGACG GACGGGCCAT GCAGGCGACA GCCGGCCCGC CGTCCGGAGG GAAGCAACAA 783  
 ATAAATCAGG GAGCGATTTG GGTGGCCTAC AATGCGTACG TCTGGATAGA GTATTTCTTT 843  
 CTTTCTTTCT TCGTGGAATA AAGTGCTCCG TGTGTGTGTG GTTGGTGGTT GTTGGTTGGA 903  
 TCAGTCAGTG TGTGTGGGTG CGTGTTGTGT ACTCAGTACT CGTGATGTGT GTGTGTGTCA 963  
 ATGTGTCAAC CCTGGTCTTC GGTGGGGGCA GCACCGAGTT GCCACCTGCC ATTCCATTTTC1023  
 CATTCCGGGC GATGAATAAA TTAAAAAGA GTCTCATTTG TTAAAAAAA AAAAAAAAAA1083  
 AA 1085

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 222 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met	Ala	Gly	Glu	Lys	Gly	Leu	Val	Leu	Leu	Asp	Phe	Trp	Val	Ser	Pro	1	5	10	15
Phe	Gly	Gln	Arg	Val	Arg	Ile	Ala	Leu	Ala	Glu	Lys	Gly	Leu	Pro	Tyr	20	25	30	
Glu	Tyr	Ala	Glu	Glu	Asp	Leu	Met	Ala	Gly	Lys	Ser	Asp	Arg	Leu	Leu	35	40	45	
Arg	Ala	Asn	Pro	Val	His	Lys	Lys	Ile	Pro	Val	Leu	Leu	His	Asp	Gly	50	55	60	
Arg	Ala	Val	Asn	Glu	Ser	Leu	Ile	Ile	Leu	Gln	Tyr	Leu	Glu	Glu	Ala	65	70	75	80
Phe	Pro	Asp	Ala	Pro	Ala	Leu	Leu	Pro	Ser	Asp	Pro	Tyr	Ala	Arg	Ala	85	90	95	
Gln	Ala	Arg	Phe	Trp	Ala	Asp	Tyr	Val	Asp	Lys	Lys	Val	Tyr	Asp	Cys	100	105	110	

Gly Ser Arg Leu Trp Lys Leu Lys Gly Glu Pro Gln Ala Gln Ala Arg  
 115 120 125  
 Ala Glu Met Leu Asp Ile Leu Lys Thr Leu Asp Gly Ala Leu Gly Asp  
 130 135 140  
 Lys Pro Phe Phe Gly Gly Asp Lys Phe Gly Phe Val Asp Ala Ala Phe  
 145 150 155 160  
 Ala Pro Phe Thr Ala Trp Phe His Ser Tyr Glu Arg Tyr Gly Glu Phe  
 165 170 175  
 Ser Leu Pro Glu Val Ala Pro Lys Ile Ala Ala Trp Ala Lys Arg Cys  
 180 185 190  
 Gly Glu Arg Glu Ser Val Ala Lys Ser Leu Tyr Ser Pro Asp Lys Val  
 195 200 205  
 Tyr Asp Phe Ile Gly Leu Leu Lys Lys Lys Tyr Gly Ile Glu  
 210 215 220

## (2) INFORMATION FOR SEQ ID NO: 3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 865 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 54..725

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 1..865
- (D) OTHER INFORMATION: /note= "WIC1 SEQUENCE AND ENCODED  
IC1 AMINO ACID SEQUENCE"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGAAGTCAAC CATTGATCTT CAAGAAGCGG AAGCAAACAG AGCAAAAGGT GTG ATG 56  
 Met

GCG GCG CCG GCG GTG AAG GTG TAC GGG TGG GCG ATG TCG CCG TTC GTG 104  
 Ala Ala Pro Ala Val Lys Val Tyr Gly Trp Ala Met Ser Pro Phe Val  
  
 GCG CGC GCG CTG CTG TGC CTG GAG GAG GCC GGC GTG GAG TAC GAG CTC 152  
 Ala Arg Ala Leu Leu Cys Leu Glu Glu Ala Gly Val Glu Tyr Glu Leu  
  
 GTC CCC ATG AGC CGC GAG GCC GGC GAC CAC CGC CAG CCC GAC TTC CTC 200  
 Val Pro Met Ser Arg Glu Ala Gly Asp His Arg Gln Pro Asp Phe Leu  
  
 GCC CGG AAC CCC TTC GGC CAG GTC CCC GTT CTC GAG GAC GGC GAC CTC 248  
 Ala Arg Asn Pro Phe Gly Gln Val Pro Val Leu Glu Asp Gly Asp Leu  
  
 ACC ATC TTC GAG TCG CGC GCC GTC GCG AGG CAC GTG CTG CGC AAG CAC 296  
 Thr Ile Phe Glu Ser Arg Ala Val Ala Arg His Val Leu Arg Lys His  
  
 AAA CCG GAG CTG CTG GGC TCC GGC TCG CCG GAG TCG GCG GCG ATG GTG 344  
 Lys Pro Glu Leu Leu Gly Ser Gly Ser Pro Glu Ser Ala Ala Met Val  
  
 GAC GTG TGG CTG GAG GTG GAG GCC CAC CAG CAC CAG ACC CCG GCG GGC 392  
 Asp Val Trp Leu Glu Val Glu Ala His Gln His Gln Thr Pro Ala Gly  
  
 ACC ATC GTC ATG CAG TGC ATC CTC ACC CCG TTC CTC GGC TGC CAG CGC 440  
 Thr Ile Val Met Gln Cys Ile Leu Thr Pro Phe Leu Gly Cys Gln Arg  
  
 GAC CAG GCC GCC ATC GAC GAG AAC GCG GCA AAG CTG ACG AAT CTG TTC 488  
 Asp Gln Ala Ala Ile Asp Glu Asn Ala Ala Lys Leu Thr Asn Leu Phe  
  
 GAC GTG TAC GAG GCG CGC CTG TCG GCG TCG AGG TAC CTT GCC GGC GAG 536  
 Asp Val Tyr Glu Ala Arg Leu Ser Ala Ser Arg Tyr Leu Ala Gly Glu  
  
 GCG GTC AGC CTC GCG GAC CTC AGC CAC TTC CCG TTC ATG CGA TAC TTC 584  
 Ala Val Ser Leu Ala Asp Leu Ser His Phe Pro Phe Met Arg Tyr Phe  
  
 ATG GAC ACC GAG TAC GCG TCG CTG GTG GAG GAG CGC CCG CAC GTG AAG 632  
 Met Asp Thr Glu Tyr Ala Ser Leu Val Glu Glu Arg Pro His Val Lys  
  
 GCG TGG TGG GAG GAG TTC AAG GCC AGC CCG GCG GCG AAG AGG GTG ACG 680  
 Ala Trp Trp Glu Glu Phe Lys Ala Ser Pro Ala Ala Lys Arg Val Thr  
  
 GAG TTC ATG CCG CCA AAC TTC GGG TTC GGA AAG AAG GCA GAG AAG 725  
 Glu Phe Met Pro Pro Asn Phe Gly Phe Gly Lys Lys Ala Glu Lys  
  
 TGATGACAAG AACGAACACC GAGCGAACAT GTTGTGTGGT CTGTGCGACC CGACCATGGC 785  
  
 TCAATGTTTT GGGCTGTTTG TGTTTCACGC ATGAATGAAT AAAACAAAAT GCTTTTGGGT 845

TTCAAAAAAA AAAAAAAAAA

865

## (2) INFORMATION FOR SEQ ID NO: 4:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 224 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```

Met Ala Ala Pro Ala Val Lys Val Tyr Gly Trp Ala Met Ser Pro Phe
 1             5             10             15

Val Ala Arg Ala Leu Leu Cys Leu Glu Glu Ala Gly Val Glu Tyr Glu
          20             25             30

Leu Val Pro Met Ser Arg Glu Ala Gly Asp His Arg Gln Pro Asp Phe
          35             40             45

Leu Ala Arg Asn Pro Phe Gly Gln Val Pro Val Leu Glu Asp Gly Asp
          50             55             60

Leu Thr Ile Phe Glu Ser Arg Ala Val Ala Arg His Val Leu Arg Lys
          65             70             75             80

His Lys Pro Glu Leu Leu Gly Ser Gly Ser Pro Glu Ser Ala Ala Met
          85             90             95

Val Asp Val Trp Leu Glu Val Glu Ala His Gln His Gln Thr Pro Ala
          100            105            110

Gly Thr Ile Val Met Gln Cys Ile Leu Thr Pro Phe Leu Gly Cys Gln
          115            120            125

Arg Asp Gln Ala Ala Ile Asp Glu Asn Ala Ala Lys Leu Thr Asn Leu
          130            135            140

Phe Asp Val Tyr Glu Ala Arg Leu Ser Ala Ser Arg Tyr Leu Ala Gly
          145            150            155            160

Glu Ala Val Ser Leu Ala Asp Leu Ser His Phe Pro Phe Met Arg Tyr
          165            170            175

```

Phe Met Asp Thr Glu Tyr Ala Ser Leu Val Glu Glu Arg Pro His Val  
                   180                  185                  190

Lys Ala Trp Trp Glu Glu Phe Lys Ala Ser Pro Ala Ala Lys Arg Val  
                   195                  200                  205

Thr Glu Phe Met Pro Pro Asn Phe Gly Phe Gly Lys Lys Ala Glu Lys  
                   210                  215                  220

## (2) INFORMATION FOR SEQ ID NO: 5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 930 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 60..725

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 1..930
- (D) OTHER INFORMATION: /note= "WIC2 SEQUENCE AND ENCODED  
 IC2 AMINO ACID SEQUENCE"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CACGCGTCCA TCTCCAAGAA GCGGAAGCTA GTGGAGCAGA GCAAACCAAG CAAGGTTGG 59

ATG GCG CCG GCG GTG AAG GTG TAC GGG TGG GCC GTG TCG CCG TTC GTG 107  
 Met Ala Pro Ala Val Lys Val Tyr Gly Trp Ala Val Ser Pro Phe Val

GCG CGC CCA CTG CTG TGC CTG GAG GAG GCC GGC GTC GAG TAC GAG CTC 155  
 Ala Arg Pro Leu Leu Cys Leu Glu Glu Ala Gly Val Glu Tyr Glu Leu

GTG TCC ATG AGC CGC GCG GCC GGC GAC CAC CGC CAG CCG GAC TTC CTC 203  
 Val Ser Met Ser Arg Ala Ala Gly Asp His Arg Gln Pro Asp Phe Leu

GCC CGG AAC CCC TTC GGC CAG GTC CCC GTC CTC GAG GAC GGC GAC CTC 251  
 Ala Arg Asn Pro Phe Gly Gln Val Pro Val Leu Glu Asp Gly Asp Leu

ACC CTC TTC GAG TCG CGC GCG ATC GCG AGG CAC GTG CTC CGG AAG CAC 299  
 Thr Leu Phe Glu Ser Arg Ala Ile Ala Arg His Val Leu Arg Lys His  
  
 AAG CCG GAG CTG CTG GGC TGC GGC TCG CCG GAG GCG GAG GCG ATG GTG 347  
 Lys Pro Glu Leu Leu Gly Cys Gly Ser Pro Glu Ala Glu Ala Met Val  
  
 GAC GTG TGG CTG GAG GTG GAG GCC CAC CAG TAC AAC CCC GCG GCC AGC 395  
 Asp Val Trp Leu Glu Val Glu Ala His Gln Tyr Asn Pro Ala Ala Ser  
  
 GCC ATC GTG GTG CAG TGC ATC ATC TTG CCG CTA CTG GGC GGC GCG CGG 443  
 Ala Ile Val Val Gln Cys Ile Ile Leu Pro Leu Leu Gly Gly Ala Arg  
  
 GAC CAG GCG GTG GTG GAC GAG AAC GTA GCC AAG CTC AAG AAG GTG CTG 491  
 Asp Gln Ala Val Val Asp Glu Asn Val Ala Lys Leu Lys Lys Val Leu  
  
 GAG GTG TAC GAG GCA CGG CTG TCG GCG TCC AGG TAC CTC GCC GGG GAC 539  
 Glu Val Tyr Glu Ala Arg Leu Ser Ala Ser Arg Tyr Leu Ala Gly Asp  
  
 GAC ATC AGC CTC GCC GAC CTC AGC CAC TTC CCC TTC ACG CGC TAC TTC 587  
 Asp Ile Ser Leu Ala Asp Leu Ser His Phe Pro Phe Thr Arg Tyr Phe  
  
 ATG GAG ACG GAG TAC GCG CCG CTG GTG GCG GAG CTC CCC CAC GTG AAC 635  
 Met Glu Thr Glu Tyr Ala Pro Leu Val Ala Glu Leu Pro His Val Asn  
  
 GCG TGG TGG GAG GGG CTC AAG GCC AGG CCG GCC GCG AGG AAG GTG ACG 683  
 Ala Trp Trp Glu Gly Leu Lys Ala Arg Pro Ala Ala Arg Lys Val Thr  
  
 GAG CTC ATG CCG CCG GAC CTT GGG CTT GGA AAG AAA GCA GAG 725  
 Glu Leu Met Pro Pro Asp Leu Gly Leu Gly Lys Lys Ala Glu  
  
 TAGTGATGAC TGCCGCCAAC GTTCACCAGG ATCGAGCAAG TCACTGTCGA GTCTCCGGTT 785  
 TTGCGTTGTA CGGCACCGGG GCACCGGCCT ATATTTTCTG TACCAGTGGC TCGTGTTTGT 845  
 ATGTTTTAGT CTCACGCTTG AATAAAATGC AAGATATACC CATCGGTTCT AAAAGAAAAA 905  
 AAAAAAAAAA AAAAAAAAAA AAAAA 930

## (2) INFORMATION FOR SEQ ID NO: 6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 222 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Ala Pro Ala Val Lys Val Tyr Gly Trp Ala Val Ser Pro Phe Val  
 1 5 10 15  
 Ala Arg Pro Leu Leu Cys Leu Glu Glu Ala Gly Val Glu Tyr Glu Leu  
 20 25 30  
 Val Ser Met Ser Arg Ala Ala Gly Asp His Arg Gln Pro Asp Phe Leu  
 35 40 45  
 Ala Arg Asn Pro Phe Gly Gln Val Pro Val Leu Glu Asp Gly Asp Leu  
 50 55 60  
 Thr Leu Phe Glu Ser Arg Ala Ile Ala Arg His Val Leu Arg Lys His  
 65 70 75 80  
 Lys Pro Glu Leu Leu Gly Cys Gly Ser Pro Glu Ala Glu Ala Met Val  
 85 90 95  
 Asp Val Trp Leu Glu Val Glu Ala His Gln Tyr Asn Pro Ala Ala Ser  
 100 105 110  
 Ala Ile Val Val Gln Cys Ile Ile Leu Pro Leu Leu Gly Gly Ala Arg  
 115 120 125  
 Asp Gln Ala Val Val Asp Glu Asn Val Ala Lys Leu Lys Lys Val Leu  
 130 135 140  
 Glu Val Tyr Glu Ala Arg Leu Ser Ala Ser Arg Tyr Leu Ala Gly Asp  
 145 150 155 160  
 Asp Ile Ser Leu Ala Asp Leu Ser His Phe Pro Phe Thr Arg Tyr Phe  
 165 170 175  
 Met Glu Thr Glu Tyr Ala Pro Leu Val Ala Glu Leu Pro His Val Asn  
 180 185 190  
 Ala Trp Trp Glu Gly Leu Lys Ala Arg Pro Ala Ala Arg Lys Val Thr  
 195 200 205  
 Glu Leu Met Pro Pro Asp Leu Gly Leu Gly Lys Lys Ala Glu  
 210 215 220

(2) INFORMATION FOR SEQ ID NO: 7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 927 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 72..707

## (ix) FEATURE:

- (A) NAME/KEY: misc feature
- (B) LOCATION: 1..927
- (D) OTHER INFORMATION: /note= "WIC 3/7/8 SEQUENCE AND ENCODED IC3 AMINO ACID SEQUENCE"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```

AGCGGCTTTA CCTACCGAGA AGAAGAGAGA AAAAAGGTTC GAGTGC GTTC CAGAGTGAGG 60
AGTGAGAAGA G ATG GCT CCG GTG AAG CTG TAC GGC GCG ACC CTG TCG TGG 110
      Met Ala Pro Val Lys Leu Tyr Gly Ala Thr Leu Ser Trp
AAC GTC ACC AGG TGC GTG GCG GCG CTG GAG GAG GCC GGC GTC CAG TAC 158
      Asn Val Thr Arg Cys Val Ala Ala Leu Glu Glu Ala Gly Val Gln Tyr
GAG ATC GTA CCC ATC AAC TTC GGC ACC GGC GAG CAC AAG AGC CCC GAC 206
      Glu Ile Val Pro Ile Asn Phe Gly Thr Gly Glu His Lys Ser Pro Asp
CAC CTC GCC AGG AAC CCC TTC GGC CAG GTG CCA GCT TTG CAG GAT GGT 254
      His Leu Ala Arg Asn Pro Phe Gly Gln Val Pro Ala Leu Gln Asp Gly
GAC TTA TAC GTC TTC GAA TCA CGT GCT ATT TGC AAG TAC GCG TGC CGC 302
      Asp Leu Tyr Val Phe Glu Ser Arg Ala Ile Cys Lys Tyr Ala Cys Arg
AAG AAC AAG CCA GAG CTG TTG AAG GAG GGC GAC ATC AAG GAG TCA GCA 350
      Lys Asn Lys Pro Glu Leu Leu Lys Glu Gly Asp Ile Lys Glu Ser Ala
ATG GTG GAT GTG TGG CTC GAG GTG GAG GCC CAT CAG TAC ACT GCC GCT 398
      Met Val Asp Val Trp Leu Glu Val Glu Ala His Gln Tyr Thr Ala Ala
CTG AGC CCC ATT CTC TTC GAG TGC CTT ATC CAT CCA ATG CTT GGG GGA 446
      Leu Ser Pro Ile Leu Phe Glu Cys Leu Ile His Pro Met Leu Gly Gly

```

GCC ACT GAC CAG AAG GTC ATC GAC GAC AAC CTT GTT AAG ATC AAG AAC 494  
 Ala Thr Asp Gln Lys Val Ile Asp Asp Asn Leu Val Lys Ile Lys Asn

GTG CTG GCG GTG TAC GAG GCG CAC CTG AGC AAG TCC AAG TAC CTG GCT 542  
 Val Leu Ala Val Tyr Glu Ala His Leu Ser Lys Ser Lys Tyr Leu Ala

GGA GAC TTC CTC AGT CTT GCG GAC CTT AAC CAT GTG TCT GTC ACC CTG 590  
 Gly Asp Phe Leu Ser Leu Ala Asp Leu Asn His Val Ser Val Thr Leu

TGC TTG GCG GCT ACA CCC TAT GCG TCT CTG TTC GAC GCG TAC CCG CAT 638  
 Cys Leu Ala Ala Thr Pro Tyr Ala Ser Leu Phe Asp Ala Tyr Pro His

GTG AAG GCC TGG TGG ACT GAC CTG CTG GCG AGG CCG TCC GTC CAG AAG 686  
 Val Lys Ala Trp Trp Thr Asp Leu Leu Ala Arg Pro Ser Val Gln Lys

GTC GCA GCG CTG ATG AAG CCA TGATCTTAAT TGCTGGTGCT CGTTCGTCGC 737  
 Val Ala Ala Leu Met Lys Pro

GAAATAAGCC GAGGTGTGTG CCCCCGATG TGTGCCTGTA CGAGTGTGTG TTCTTGTGAT 797

GTCTCCTCGT GTTGAATGTT CAGGCTTGTG CTTGCGATCC TGTCTCATCT TTTACTGAAA 857

TGAGCGTTCC TATGCTCTGG TTTAATAATA AATTGTGCCT AGATATTATC TCAAAAAAAAA 917

AAAAAAAAAA 927

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 212 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met	Ala	Pro	Val	Lys	Leu	Tyr	Gly	Ala	Thr	Leu	Ser	Trp	Asn	Val	Thr
1				5					10					15	
Arg	Cys	Val	Ala	Ala	Leu	Glu	Glu	Ala	Gly	Val	Gln	Tyr	Glu	Ile	Val
			20					25					30		
Pro	Ile	Asn	Phe	Gly	Thr	Gly	Glu	His	Lys	Ser	Pro	Asp	His	Leu	Ala
			35				40					45			

Arg Asn Pro Phe Gly Gln Val Pro Ala Leu Gln Asp Gly Asp Leu Tyr  
 50 55 60  
 Val Phe Glu Ser Arg Ala Ile Cys Lys Tyr Ala Cys Arg Lys Asn Lys  
 65 70 75 80  
 Pro Glu Leu Leu Lys Glu Gly Asp Ile Lys Glu Ser Ala Met Val Asp  
 85 90 95  
 Val Trp Leu Glu Val Glu Ala His Gln Tyr Thr Ala Ala Leu Ser Pro  
 100 105 110  
 Ile Leu Phe Glu Cys Leu Ile His Pro Met Leu Gly Gly Ala Thr Asp  
 115 120 125  
 Gln Lys Val Ile Asp Asp Asn Leu Val Lys Ile Lys Asn Val Leu Ala  
 130 135 140  
 Val Tyr Glu Ala His Leu Ser Lys Ser Lys Tyr Leu Ala Gly Asp Phe  
 145 150 155 160  
 Leu Ser Leu Ala Asp Leu Asn His Val Ser Val Thr Leu Cys Leu Ala  
 165 170 175  
 Ala Thr Pro Tyr Ala Ser Leu Phe Asp Ala Tyr Pro His Val Lys Ala  
 180 185 190  
 Trp Trp Thr Asp Leu Leu Ala Arg Pro Ser Val Gln Lys Val Ala Ala  
 195 200 205  
 Leu Met Lys Pro  
 210

## (2) INFORMATION FOR SEQ ID NO: 9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 866 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS

(B) LOCATION:45..683

## (ix) FEATURE:

(A) NAME/KEY: misc\_feature

(B) LOCATION:1..866

(D) OTHER INFORMATION:/note= "WIC5 SEQUENCE AND ENCODED  
IC5 AMINO ACID SEQUENCE"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GAAGCAGGCA ACAGGCGAGC AGGAAGGAAG CAAGAGAGGT GGAG ATG GCG CCC ATC	56
Met Ala Pro Ile	
AAG CTG TAC GGG ATG ATG CTG TCG GCC AAC GTG ACC CGC GTG ACC ACG	104
Lys Leu Tyr Gly Met Met Leu Ser Ala Asn Val Thr Arg Val Thr Thr	
CTG CTC AAC GAG CTC GGC CTC GAG TTC GAC TTC GTC GAC GTC GAC CTC	152
Leu Leu Asn Glu Leu Gly Leu Glu Phe Asp Phe Val Asp Val Asp Leu	
CGC ACC GGC GCC CAC AAG CAC CCC GAC TTC CTC AAG CTC AAC CCT TTC	200
Arg Thr Gly Ala His Lys His Pro Asp Phe Leu Lys Leu Asn Pro Phe	
GGC CAG ATC CCC GCG CTG CAG GAC GGA GAC GAA GTT GTC TTC GAG TCG	248
Gly Gln Ile Pro Ala Leu Gln Asp Gly Asp Glu Val Val Phe Glu Ser	
CGC GCC ATC AAC CGG TAC ATC GCG ACC AAG TAC GGG GCG TCC CTG CTG	296
Arg Ala Ile Asn Arg Tyr Ile Ala Thr Lys Tyr Gly Ala Ser Leu Leu	
CCG ACG CCG TCG GCC AAG CTG GAG GCG TGG CTG GAG GTG GAG TCG CAC	344
Pro Thr Pro Ser Ala Lys Leu Glu Ala Trp Leu Glu Val Glu Ser His	
CAC TTC TAC CCG CCG GCG CGG ACG CTG GTG TAC GAG CTG GTC ATC AAG	392
His Phe Tyr Pro Pro Ala Arg Thr Leu Val Tyr Glu Leu Val Ile Lys	
CCC ATG CTG GGC GCC CCC ACC GAC GCC GCC GAG GTG GAC AAG AAC GCC	440
Pro Met Leu Gly Ala Pro Thr Asp Ala Ala Glu Val Asp Lys Asn Ala	
GCC GAC CTC GCC AAG CTG CTC GAC GTC TAC GAG GCC CAC CTC GCC GCC	488
Ala Asp Leu Ala Lys Leu Leu Asp Val Tyr Glu Ala His Leu Ala Ala	
GGG AAC AAG TAC CTG GCC GGC GAC GCC TTC CCG CTC GCC GAC GCC AAC	536
Gly Asn Lys Tyr Leu Ala Gly Asp Ala Phe Pro Leu Ala Asp Ala Asn	
CAC ATG TCC TAC CTC TTC ATG CTC ACC AAG AGC CCC AAG GCG GAC CTG	584
His Met Ser Tyr Leu Phe Met Leu Thr Lys Ser Pro Lys Ala Asp Leu	

GTG GCC TCC CGC CCG CAC GTC AAG GCC TGG TGG GAG GAG ATC TCC GCC 632  
Val Ala Ser Arg Pro His Val Lys Ala Trp Trp Glu Glu Ile Ser Ala

CGC CCC GCC TGG GCC AAG ACC GTC GCC TCC ATC CCC CTC CCG CCC GCC 680  
Arg Pro Ala Trp Ala Lys Thr Val Ala Ser Ile Pro Leu Pro Pro Ala

GTC TGAGGTTGCT TGTTTGGCTG CGGCGAGAAC GGAATAAAAT CGCGATGATG 733  
Val

GAATAACAA CTTTTAGAG AGGAAGCTTG GAATTCTTGG TGTTGCTGCT GTTGAATGTT 793

GAATCTTGGT GTTGAATGTT TACGGCACAT CTAATTTATC CAGTTTTTTT GCGTGAAAA 853

AAAAAAAAAA AAA 866

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 213 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met Ala Pro Ile Lys Leu Tyr Gly Met Met Leu Ser Ala Asn Val Thr  
1 5 10 15

Arg Val Thr Thr Leu Leu Asn Glu Leu Gly Leu Glu Phe Asp Phe Val  
20 25 30

Asp Val Asp Leu Arg Thr Gly Ala His Lys His Pro Asp Phe Leu Lys  
35 40 45

Leu Asn Pro Phe Gly Gln Ile Pro Ala Leu Gln Asp Gly Asp Glu Val  
50 55 60

Val Phe Glu Ser Arg Ala Ile Asn Arg Tyr Ile Ala Thr Lys Tyr Gly  
65 70 75 80

Ala Ser Leu Leu Pro Thr Pro Ser Ala Lys Leu Glu Ala Trp Leu Glu  
85 90 95

Val Glu Ser His His Phe Tyr Pro Pro Ala Arg Thr Leu Val Tyr Glu  
100 105 110

Leu Val Ile Lys Pro Met Leu Gly Ala Pro Thr Asp Ala Ala Glu Val  
 115 120 125

Asp Lys Asn Ala Ala Asp Leu Ala Lys Leu Leu Asp Val Tyr Glu Ala  
 130 135 140

His Leu Ala Ala Gly Asn Lys Tyr Leu Ala Gly Asp Ala Phe Pro Leu  
 145 150 155 160

Ala Asp Ala Asn His Met Ser Tyr Leu Phe Met Leu Thr Lys Ser Pro  
 165 170 175

Lys Ala Asp Leu Val Ala Ser Arg Pro His Val Lys Ala Trp Trp Glu  
 180 185 190

Glu Ile Ser Ala Arg Pro Ala Trp Ala Lys Thr Val Ala Ser Ile Pro  
 195 200 205

Leu Pro Pro Ala Val  
 210

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 897 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 15..668

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 1..897
- (D) OTHER INFORMATION: /note= "WIC4 SEQUENCE AND ENCODED  
 IC4 AMINO ACID SEQUENCE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

AACCAAGGGA AACA ATG GCG CCG GTG AAG GTG TTC GGG CCG GCG ATG TCG 50

Met Ala Pro Val Lys Val Phe Gly Pro Ala Met Ser

ACC AAC GTG GCC CGG GTG CTG GTG TGC CTG GAG GAG GTC GGC GCC GAG 98  
 Thr Asn Val Ala Arg Val Leu Val Cys Leu Glu Glu Val Gly Ala Glu

TAC GAG GTG GTC GAC ATC GAT TTC AAG GCC ATG GAG CAC AAG AGC CCC 146  
 Tyr Glu Val Val Asp Ile Asp Phe Lys Ala Met Glu His Lys Ser Pro

GAG CAT CTC GTC AGA AAC CCG TTC GGC CAA ATC CCT GCC TTC CAG GAT 194  
 Glu His Leu Val Arg Asn Pro Phe Gly Gln Ile Pro Ala Phe Gln Asp

GGG GAT CTG CTT CTC TTC GAG TCA CGC GCA ATT GCG AGG TAC GTG CTC 242  
 Gly Asp Leu Leu Leu Phe Glu Ser Arg Ala Ile Ala Arg Tyr Val Leu

CGC AAG TAC AAG AAG AAC GAA GTG GAC CTG CTG AGG GAA GGC GAC CTC 290  
 Arg Lys Tyr Lys Lys Asn Glu Val Asp Leu Leu Arg Glu Gly Asp Leu

AAG GAG GCG GCG ATG GTG GAC GTA TGG ACG GAG GTG GAC GCG CAC ACC 338  
 Lys Glu Ala Ala Met Val Asp Val Trp Thr Glu Val Asp Ala His Thr

TAC AAC CCG GCC ATC TCG CCG ATC GTG TAC GAG TGC TCA TCA ACC GCT 386  
 Tyr Asn Pro Ala Ile Ser Pro Ile Val Tyr Glu Cys Ser Ser Thr Ala

CAT GCG CGG CTG CCG ACC AAC CAA ACG GTG GTG GAC GAG AGC CTG GAG 434  
 His Ala Arg Leu Pro Thr Asn Gln Thr Val Val Asp Glu Ser Leu Glu

AAG CTC AAG AAC GTG CTG GAG GTC TAC GAG GCG CGC CTG TCC AAG CAC 482  
 Lys Leu Lys Asn Val Leu Glu Val Tyr Glu Ala Arg Leu Ser Lys His

GAC TAC CTC GCC GGG GAC TTC GTC AGC TTC GCG GAC CTC AAC CAC TTC 530  
 Asp Tyr Leu Ala Gly Asp Phe Val Ser Phe Ala Asp Leu Asn His Phe

CCC TAC ACC TTC TAC TTC ATG GCC ACG CCG CAC GCG GCC CTC TTC GAC 578  
 Pro Tyr Thr Phe Tyr Phe Met Ala Thr Pro His Ala Ala Leu Phe Asp

TCG TAC CCG CAC GTC AAG GCC TGG TGG GAG AGG ATC ATG GCG AGG CCG 626  
 Ser Tyr Pro His Val Lys Ala Trp Trp Glu Arg Ile Met Ala Arg Pro

GCC GTG AAG AAG CTC GCC GCG CAG ATG GTT CCC AAG AAG CCG 668  
 Ala Val Lys Lys Leu Ala Ala Gln Met Val Pro Lys Lys Pro

TGATTTGCTA GGC GGGATCT CGCATCGTGG GATCCGATTC CGATCACTGA TCTGTGTGGC 728

GTTTTCTTTT CTGTGTGGTG TCGCAATAA GGCAAATGAG CTCGTGTGTG TGTGGCTGGA 788

ATTGCACCAG CGTGCAGTTT TTGCGCTTTG CGTGTGTGTG GTCGTGAAAA CTCTTGAGAT 848

GGAACAATGT CTTCGTAATG CTTTCACATT TTAAAAAAAA AAAAAAAAAA

897

## (2) INFORMATION FOR SEQ ID NO: 12:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 218 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Ala Pro Val Lys Val Phe Gly Pro Ala Met Ser Thr Asn Val Ala  
 1 5 10 15

Arg Val Leu Val Cys Leu Glu Glu Val Gly Ala Glu Tyr Glu Val Val  
 20 25 30

Asp Ile Asp Phe Lys Ala Met Glu His Lys Ser Pro Glu His Leu Val  
 35 40 45

Arg Asn Pro Phe Gly Gln Ile Pro Ala Phe Gln Asp Gly Asp Leu Leu  
 50 55 60

Leu Phe Glu Ser Arg Ala Ile Ala Arg Tyr Val Leu Arg Lys Tyr Lys  
 65 70 75 80

Lys Asn Glu Val Asp Leu Leu Arg Glu Gly Asp Leu Lys Glu Ala Ala  
 85 90 95

Met Val Asp Val Trp Thr Glu Val Asp Ala His Thr Tyr Asn Pro Ala  
 100 105 110

Ile Ser Pro Ile Val Tyr Glu Cys Ser Ser Thr Ala His Ala Arg Leu  
 115 120 125

Pro Thr Asn Gln Thr Val Val Asp Glu Ser Leu Glu Lys Leu Lys Asn  
 130 135 140

Val Leu Glu Val Tyr Glu Ala Arg Leu Ser Lys His Asp Tyr Leu Ala  
 145 150 155 160

Gly Asp Phe Val Ser Phe Ala Asp Leu Asn His Phe Pro Tyr Thr Phe  
 165 170 175

Tyr Phe Met Ala Thr Pro His Ala Ala Leu Phe Asp Ser Tyr Pro His

180	185	190
Val Lys Ala Trp Trp Glu Arg Ile Met Ala Arg Pro Ala Val Lys Lys		
195	200	205
Leu Ala Ala Gln Met Val Pro Lys Lys Pro		
210	215	

## (2) INFORMATION FOR SEQ ID NO: 13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 721 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 21..686

## (ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 1..721
- (D) OTHER INFORMATION: /note= "TA 27 SEQUENCE AND ENCODED AMINO ACID SEQUENCE"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TTCGGCACGA GGAAGAAGGG ATG GAG CCT ATG AAG GTG TAC GGC TGG GCG	50
Met Glu Pro Met Lys Val Tyr Gly Trp Ala	
GTG TCG CCA TGG ATG GCG CGG GTC CTC GTC TCC CTG GAG GAG GCC GGC	98
Val Ser Pro Trp Met Ala Arg Val Leu Val Ser Leu Glu Glu Ala Gly	
GCC GAC TAC GAG CTC GTG CCC ATG AGC CGC AAC GGC GGC GAC CAC CGG	146
Ala Asp Tyr Glu Leu Val Pro Met Ser Arg Asn Gly Gly Asp His Arg	
CGG CCG GAG CAC CTC GCC AGA AAC CCC TTC GGT GAG ATC CCG GTG CTC	194
Arg Pro Glu His Leu Ala Arg Asn Pro Phe Gly Glu Ile Pro Val Leu	
GAA TAC GGC GGT CTG ACG CTT TAC CAA TCC CGC GCC ATT GCA AGG CAT	242
Glu Tyr Gly Gly Leu Thr Leu Tyr Gln Ser Arg Ala Ile Ala Arg His	

ATT CTC CGC AAA CAC AAG CCC GGG CTT CTA GGA GCA GGC AGC CTC GAG 290  
 Ile Leu Arg Lys His Lys Pro Gly Leu Leu Gly Ala Gly Ser Leu Glu  
  
 GAG TCG GCG ATG GTG GAT GTA TGG GTC GAC GTG GAT GCC CAC CAC CTG 338  
 Glu Ser Ala Met Val Asp Val Trp Val Asp Val Asp Ala His His Leu  
  
 GAG CCC GTA CTC AAG CCC ATC GTG TGG AAC TGC ATC ATC AAC CCG TTC 386  
 Glu Pro Val Leu Lys Pro Ile Val Trp Asn Cys Ile Ile Asn Pro Phe  
  
 GTC GGG AGG GAC GTC GAC CAG GGC CTC GTC GAT GAG AGC GTC GAG AAG 434  
 Val Gly Arg Asp Val Asp Gln Gly Leu Val Asp Glu Ser Val Glu Lys  
  
 CTC AAG AAG CTG CTG GAG GTG TAC GAG GCA AGA CTG TCA AGC AAC AAG 482  
 Leu Lys Lys Leu Leu Glu Val Tyr Glu Ala Arg Leu Ser Ser Asn Lys  
  
 TAC TTG GCC GGG GAT TTC GTC AGC TTC GCC GAC CTC ACC CAT TTC TCC 530  
 Tyr Leu Ala Gly Asp Phe Val Ser Phe Ala Asp Leu Thr His Phe Ser  
  
 TTC ATG CGC TAC TTC ATG GCG ACG GAG CAT GCG GTT GTG CTC GAT GCG 578  
 Phe Met Arg Tyr Phe Met Ala Thr Glu His Ala Val Val Leu Asp Ala  
  
 TAT CCG CAT GTG AAG GCA TGG TGG AAG GCG CTG CTG GCA AGG CCA TCG 626  
 Tyr Pro His Val Lys Ala Trp Trp Lys Ala Leu Leu Ala Arg Pro Ser  
  
 GTC AAG AAG GTG ATA GCT GGC ATG CCT CCG GAT TTT GGA TTC GGG AGC 674  
 Val Lys Lys Val Ile Ala Gly Met Pro Pro Asp Phe Gly Phe Gly Ser  
  
 GGG AGA ATA CCA TGATAAAGCA TGCTTGTTTG TCTATGATGC TCTGA 721  
 Gly Arg Ile Pro

## (2) INFORMATION FOR SEQ ID NO: 14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 222 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met Glu Pro Met Lys Val Tyr Gly Trp Ala Val Ser Pro Trp Met Ala  
 1 5 10 15  
 Arg Val Leu Val Ser Leu Glu Glu Ala Gly Ala Asp Tyr Glu Leu Val  
 20 25 30

Pro Met Ser Arg Asn Gly Gly Asp His Arg Arg Pro Glu His Leu Ala  
           35                          40                          45  
 Arg Asn Pro Phe Gly Glu Ile Pro Val Leu Glu Tyr Gly Gly Leu Thr  
           50                          55                          60  
 Leu Tyr Gln Ser Arg Ala Ile Ala Arg His Ile Leu Arg Lys His Lys  
           65                          70                          75                          80  
 Pro Gly Leu Leu Gly Ala Gly Ser Leu Glu Glu Ser Ala Met Val Asp  
                           85                          90                          95  
 Val Trp Val Asp Val Asp Ala His His Leu Glu Pro Val Leu Lys Pro  
                           100                          105                          110  
 Ile Val Trp Asn Cys Ile Ile Asn Pro Phe Val Gly Arg Asp Val Asp  
           115                          120                          125  
 Gln Gly Leu Val Asp Glu Ser Val Glu Lys Leu Lys Lys Leu Leu Glu  
           130                          135                          140  
 Val Tyr Glu Ala Arg Leu Ser Ser Asn Lys Tyr Leu Ala Gly Asp Phe  
           145                          150                          155                          160  
 Val Ser Phe Ala Asp Leu Thr His Phe Ser Phe Met Arg Tyr Phe Met  
                           165                          170                          175  
 Ala Thr Glu His Ala Val Val Leu Asp Ala Tyr Pro His Val Lys Ala  
           180                          185                          190  
 Trp Trp Lys Ala Leu Leu Ala Arg Pro-Ser Val Lys Lys Val Ile Ala  
           195                          200                          205  
 Gly Met Pro Pro Asp Phe Gly Phe Gly Ser Gly Arg Ile Pro  
           210                          215                          220

## (2) INFORMATION FOR SEQ ID NO: 15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 926 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:66..764

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

```

AACCACTTTC ATCAACGTCT CCTACGCTCA CCGTTCGTTG CTCCGCACAT CAGCAGGACT 60
TGCCA ATG GCG GGA GAC GGC GAG CTG AAG CTG CTG GGC GTG TGG ACG 107
  Met Ala Gly Asp Gly Glu Leu Lys Leu Leu Gly Val Trp Thr
    1          5          10

AGC CCG TTC GTC ATC AGG GTG CGC GTG GTG CTC AAC CTC AAG TCG CTG 155
Ser Pro Phe Val Ile Arg Val Arg Val Val Leu Asn Leu Lys Ser Leu
  15          20          25          30

CCG TAC GAG TAC GTG GAG GAG AGC CTG GGC AGC AAG AGC GCG CTC CTC 203
Pro Tyr Glu Tyr Val Glu Glu Ser Leu Gly Ser Lys Ser Ala Leu Leu
          35          40          45

CTG GGC TCC AAC CCG GTG CAC CAG AGC GTG CCC GTC CTC CTC CAC GGC 251
Leu Gly Ser Asn Pro Val His Gln Ser Val Pro Val Leu Leu His Gly
          50          55          60

GGC CGC CCC GTG AAC GAG TCC CAG GTC ATC GTG CAG TAC ATC GAC GAG 299
Gly Arg Pro Val Asn Glu Ser Gln Val Ile Val Gln Tyr Ile Asp Glu
          65          70          75

GTC TGG GCG GGG GCC GGC CCG TCC GTG CTC CCG GCC GAC CCC TAC GAG 347
Val Trp Ala Gly Ala Gly Pro Ser Val Leu Pro Ala Asp Pro Tyr Glu
          80          85          90

CGC GCC ACG GCG CGC TTC TGG GCG GCG TAC GTC GAC GAC AAG GTC GGG 395
Arg Ala Thr Ala Arg Phe Trp Ala Ala Tyr Val Asp Asp Lys Val Gly
          95          100          105          110

TCG GCG TGG ACG GGG ATG CTC TTC TCG TGC AAG ACG GAG GAG GAG CGG 443
Ser Ala Trp Thr Gly Met Leu Phe Ser Cys Lys Thr Glu Glu Glu Arg
          115          120          125

GCG GAG GCG GTG TCC CGG GCC GTG GCG GCG CTG GAG ACC CTG GAG GGC 491
Ala Glu Ala Val Ser Arg Ala Val Ala Ala Leu Glu Thr Leu Glu Gly
          130          135          140

```

GCG TTC GCG GAG TGC TCC AAG GGG AAG GCG TTC TTC GGC GGC GAC GCC 539  
Ala Phe Ala Glu Cys Ser Lys Gly Lys Ala Phe Phe Gly Gly Asp Ala  
145 150 155

ATC GGG TTC GTC GAC GTC GTG CTT GGC GGC TAC CTC GGC TGG TTC GGC 587  
Ile Gly Phe Val Asp Val Val Leu Gly Gly Tyr Leu Gly Trp Phe Gly  
160 165 170

GCG ATC GAC AAG ATC ATC GGG CGC CGG CTG ATC GAC CCG GCG AGG ACG 635  
Ala Ile Asp Lys Ile Ile Gly Arg Arg Leu Ile Asp Pro Ala Arg Thr  
175 180 185 190

CCG CTG CTG GCC AGG TGG GAG GAG CGG TTC CGC GCG GCG GAC GCG GCC 683  
Pro Leu Leu Ala Arg Trp Glu Glu Arg Phe Arg Ala Ala Asp Ala Ala  
195 200 205

AAG GGC GTC GTG CCG GAC GAC GCC GAC AAG ATG CTC GAG TTC TTG CCC 731  
Lys Gly Val Val Pro Asp Asp Ala Asp Lys Met Leu Glu Phe Leu Pro  
210 215 220

ACC GTG CTC GCT TGG ATC GCC GGC AAA GCG AAG TGA ACTGTGT CTGTGAGGCC 784  
Thr Val Leu Ala Trp Ile Ala Gly Lys Ala Lys  
225 230

GTGACATCGC CAGCTCGTGA CATGTGTGTT TGTGTGTGTC TGAGTCCGTC CAGTGTGTGC 844

TGAATAAATG CACCGCATGT CGTGTGTTGT ACCAAGGGCA AACAAATGCTG AATAATTTTG 904

CTGTTAAAAA AAAAAAAAAA·AA 926

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 233 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Ala Gly Asp Gly Glu Leu Lys Leu Leu Gly Val Trp Thr Ser Pro  
1 5 10 15

Phe Val Ile Arg Val Arg Val Val Leu Asn Leu Lys Ser Leu Pro Tyr  
20 25 30

Glu Tyr Val Glu Glu Ser Leu Gly Ser Lys Ser Ala Leu Leu Gly  
           35                          40                          45  
 Ser Asn Pro Val His Gln Ser Val Pro Val Leu Leu His Gly Gly Arg  
           50                          55                          60  
 Pro Val Asn Glu Ser Gln Val Ile Val Gln Tyr Ile Asp Glu Val Trp  
           65                          70                          75                          80  
 Ala Gly Ala Gly Pro Ser Val Leu Pro Ala Asp Pro Tyr Glu Arg Ala  
                           85                          90                          95  
 Thr Ala Arg Phe Trp Ala Ala Tyr Val Asp Asp Lys Val Gly Ser Ala  
                           100                          105                          110  
 Trp Thr Gly Met Leu Phe Ser Cys Lys Thr Glu Glu Glu Arg Ala Glu  
                           115                          120                          125  
 Ala Val Ser Arg Ala Val Ala Ala Leu Glu Thr Leu Glu Gly Ala Phe  
                           130                          135                          140  
 Ala Glu Cys Ser Lys Gly Lys Ala Phe Phe Gly Gly Asp Ala Ile Gly  
           145                          150                          155                          160  
 Phe Val Asp Val Val Leu Gly Gly Tyr Leu Gly Trp Phe Gly Ala Ile  
                           165                          170                          175  
 Asp Lys Ile Ile Gly Arg Arg Leu Ile Asp Pro Ala Arg Thr Pro Leu  
                           180                          185                          190  
 Leu Ala Arg Trp Glu Glu Arg Phe Arg Ala Ala Asp Ala Ala Lys Gly  
                           195                          200                          205  
 Val Val Pro Asp Asp Ala Asp Lys Met Leu Glu Phe Leu Pro Thr Val  
           210                          215                          220  
 Leu Ala Trp Ile Ala Gly Lys Ala Lys  
           225                          230

## (2) INFORMATION FOR SEQ ID NO: 17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1043 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:39..767

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

AGGACACGAG TATCAGGGAG GAAGACGAGG AAACGTTG ATG GCC GGC GGT GAA	53
Met Ala Gly Gly Glu	
235	
GAG CTG AAG CTG CTG GGG TGG TGG GCG CCC GGG GTG AGT CCC TAC GTG	101
Glu Leu Lys Leu Leu Gly Trp Trp Ala Pro Gly Val Ser Pro Tyr Val	
240 245 250	
CTG CGC GCC CAG ATG GCG CTC GCC GTA AAG GGG CTG AGC TAC GAC TAC	149
Leu Arg Ala Gln Met Ala Leu Ala Val Lys Gly Leu Ser Tyr Asp Tyr	
255 260 265 270	
CTC CCC GAG GAC CGC TGG TCC ACG AGC GAC CTC CTC ATC GCG TCC AAC	197
Leu Pro Glu Asp Arg Trp Ser Thr Ser Asp Leu Leu Ile Ala Ser Asn	
275 280 285	
CCC GTG TAC AAG AAG GTG CCC GTC CTC ATT CAC AAC GGC AGG CCC GTC	245
Pro Val Tyr Lys Lys Val Pro Val Leu Ile His Asn Gly Arg Pro Val	
290 295 300	
TGC GAG TCG CTG CTC ATC CTG GAG TAC CTC GAC GAC GCC GTC GGC CTT	293
Cys Glu Ser Leu Leu Ile Leu Glu Tyr Leu Asp Asp Ala Val Gly Leu	
305 310 315	
GCC GGC AAC GGC AAG CCC ATC CTC CCC GCA GAC CCC TAC AGC CGC GCC	341
Ala Gly Asn Gly Lys Pro Ile Leu Pro Ala Asp Pro Tyr Ser Arg Ala	
320 325 330	
GTC GCT CGC TTC TGG GCC GCC TAT GTG AAC GAC AAG CTG TTC CCT TCG	389
Val Ala Arg Phe Trp Ala Ala Tyr Val Asn Asp Lys Leu Phe Pro Ser	
335 340 345 350	
TGC ACC GGG ATC CTC AAG ACT ACG AAG CAG GAG GAG AGA GCC GGT AAG	437
Cys Thr Gly Ile Leu Lys Thr Thr Lys Gln Glu Glu Arg Ala Gly Lys	
355 360 365	
ATG GAG GAG ACC CTG TCC GGG CTC AGA CAC TTA GAA GCT GTC ATG GCG	485

Met Glu Glu Thr Leu Ser Gly Leu Arg His Leu Glu Ala Val Met Ala  
 370 375 380

GAG TGC TCC GAA GGG GAG GCG GAG GCG CCG TTC TTC GGT GGT GAC GCC 533  
 Glu Cys Ser Glu Gly Glu Ala Glu Ala Pro Phe Phe Gly Gly Asp Ala  
 385 390 395

ATC GGG TTC CTC GAC ATC GCG CTC GGG TGC TAT CTT CCC TGG TTT GAG 581  
 Ile Gly Phe Leu Asp Ile Ala Leu Gly Cys Tyr Leu Pro Trp Phe Glu  
 400 405 410

GCA GCA GGC CGC CTG GCC GGC TTG GGG CCG ATC ATC GAC CCG GCG AGG 629  
 Ala Ala Gly Arg Leu Ala Gly Leu Gly Pro Ile Ile Asp Pro Ala Arg  
 415 420 425 430

ACG CCG AAA CTA GCT GCG TGG GCG GAG CGG TTC AGC GTC GCC GAG CCG 677  
 Thr Pro Lys Leu Ala Ala Trp Ala Glu Arg Phe Ser Val Ala Glu Pro  
 435 440 445

ATC AAG GCG CTG CTG CCT GGG GTC GAC AAG CTG GAG GAG TAC ATC ACT 725  
 Ile Lys Ala Leu Leu Pro Gly Val Asp Lys Leu Glu Glu Tyr Ile Thr  
 450 455 460

ACG GCG CTT TAT CCA AAG TGG AAC ATC GCG GTC ACC GGC AAC 767  
 Thr Ala Leu Tyr Pro Lys Trp Asn Ile Ala Val Thr Gly Asn  
 465 470 475

TAATTAAAGA TCTTGTCGTT CCACTATGGC AAAAGAAATA AAAAAGGGCG TCGTTCGATA 827

ACCGGCGGAG GATCTCTGCC TTGTGAGTAG CTGTTTTTAC GTCAAGAGTT GAACTGTTAC 887

TACTAAGTCG GGTTCCTTTT TGCAGGGGTT AGTGGGTCGT GGTCATGAAT AATGCACAGG 947

CGTGCACTCT CTTGATCTG AGTTGTGATA TGTTGTTTCG TGAATAAATT GAAGCGTCGT1007

CGATCTTGCA TCTAAAAAAA AAAAAAAAAA AAAAAA 1043

## (2) INFORMATION FOR SEQ ID NO: 18:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 243 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Ala Gly Gly Glu Glu Leu Lys Leu Leu Gly Trp Trp Ala Pro Gly  
 1 5 10 15  
 Val Ser Pro Tyr Val Leu Arg Ala Gln Met Ala Leu Ala Val Lys Gly  
 20 25 30  
 Leu Ser Tyr Asp Tyr Leu Pro Glu Asp Arg Trp Ser Thr Ser Asp Leu  
 35 40 45  
 Leu Ile Ala Ser Asn Pro Val Tyr Lys Lys Val Pro Val Leu Ile His  
 50 55 60  
 Asn Gly Arg Pro Val Cys Glu Ser Leu Leu Ile Leu Glu Tyr Leu Asp  
 65 70 75 80  
 Asp Ala Val Gly Leu Ala Gly Asn Gly Lys Pro Ile Leu Pro Ala Asp  
 85 90 95  
 Pro Tyr Ser Arg Ala Val Ala Arg Phe Trp Ala Ala Tyr Val Asn Asp  
 100 105 110  
 Lys Leu Phe Pro Ser Cys Thr Gly Ile Leu Lys Thr Thr Lys Gln Glu  
 115 120 125  
 Glu Arg Ala Gly Lys Met Glu Glu Thr Leu Ser Gly Leu Arg His Leu  
 130 135 140  
 Glu Ala Val Met Ala Glu Cys Ser Glu Gly Glu Ala Glu Ala Pro Phe  
 145 150 155 160  
 Phe Gly Gly Asp Ala Ile Gly Phe Leu Asp Ile Ala Leu Gly Cys Tyr  
 165 170 175  
 Leu Pro Trp Phe Glu Ala Ala Gly Arg Leu Ala Gly Leu Gly Pro Ile  
 180 185 190  
 Ile Asp Pro Ala Arg Thr Pro Lys Leu Ala Ala Trp Ala Glu Arg Phe  
 195 200 205  
 Ser Val Ala Glu Pro Ile Lys Ala Leu Leu Pro Gly Val Asp Lys Leu  
 210 215 220  
 Glu Glu Tyr Ile Thr Thr Ala Leu Tyr Pro Lys Trp Asn Ile Ala Val  
 225 230 235 240  
 Thr Gly Asn

## (2) INFORMATION FOR SEQ ID NO: 19:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "primer"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

AGGTAGTTAC ATATGGCCGG AGGA

24